

0590
2107

#2

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/988,292

DATE: 01/15/2002

TIME: 18:58:37

Input Set : N:\Crf3\RULE60\09988292.raw

Output Set: N:\CRF3\01152002\I988292.raw

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Yu, Guo-Liang

6 Rosen, Craig

8 (ii) TITLE OF INVENTION: Colon Specific Genes and Proteins

10 (iii) NUMBER OF SEQUENCES: 24

12 (iv) CORRESPONDENCE ADDRESS:

13 (A) ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
14 Stewart & Olstein

15 (B) STREET: 6 Becker Farm Road

16 (C) CITY: Roseland

17 (D) STATE: NJ

18 (E) COUNTRY: USA

19 (F) ZIP: 07068-1739

21 (v) COMPUTER READABLE FORM:

22 (A) MEDIUM TYPE: Floppy disk

23 (B) COMPUTER: IBM PC compatible

24 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

25 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

27 (vi) CURRENT APPLICATION DATA:

C--> 28 (A) APPLICATION NUMBER: US/09/988,292

C--> 29 (B) FILING DATE: 19-Nov-2001

30 (C) CLASSIFICATION:

32 (vii) PRIOR APPLICATION DATA:

33 (A) APPLICATION NUMBER: 09/224,110

34 (B) FILING DATE:

36 (viii) ATTORNEY/AGENT INFORMATION:

37 (A) NAME: Ferraro, Gregory D.

38 (B) REGISTRATION NUMBER: 36,134

39 (C) REFERENCE/DOCKET NUMBER: 325800-435

41 (ix) TELECOMMUNICATION INFORMATION:

42 (A) TELEPHONE: 201-994-1700

43 (B) TELEFAX: 201-994-1744

46 (2) INFORMATION FOR SEQ ID NO: 1:

48 (i) SEQUENCE CHARACTERISTICS:

49 (A) LENGTH: 638 base pairs

50 (B) TYPE: nucleic acid

51 (C) STRANDEDNESS: single

52 (D) TOPOLOGY: linear

54 (ii) MOLECULE TYPE: cDNA

57 (ix) FEATURE:

58 (A) NAME/KEY: mat_peptide

59 (B) LOCATION: 1..501

61 (ix) FEATURE:

62 (A) NAME/KEY: CDS

63 (B) LOCATION: 1..501

66 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/988,292

DATE: 01/15/2002

TIME: 18:58:37

Input Set : N:\Crf3\RULE60\09988292.raw

Output Set: N:\CRF3\01152002\I988292.raw

```

68 GCC AGG CAG CTG GCT GCC SAC CAG GCC GTG TAT GTG AAG GTC AAG GCT      48
W--> 69 Ala Arg Gln Leu Ala Ala Xaa Gln Ala Val Tyr Val Lys Val Lys Ala
70   1           5           10           15
72 GAA GCC CGG GAA CTG CTG GGC CAC CCG TGG TCT CTG TGT CCT GTC TGT      96
73 Glu Ala Arg Glu Leu Leu Gly His Pro Trp Ser Leu Cys Pro Val Cys
74           20           25           30
76 GGG TGC CAA CTC ACC ACC TTT GAT GGG GCC CGT GGT GCC ACC ACT CTC      144
77 Gly Cys Gln Leu Thr Thr Phe Asp Gly Ala Arg Gly Ala Thr Thr Leu
78           35           40           45
80 CTG GTG TCT ATG AAG CTC TCT TCC CGC TGC CCA GGA CTA CAG AAT ACC      192
81 Leu Val Ser Met Lys Leu Ser Ser Arg Cys Pro Gly Leu Gln Asn Thr
82           50           55           60
84 ATC CCC TGG TAC CGT GTA GTT GCC GAA GTC CAG ATC TGC CAT GGC AAA      240
85 Ile Pro Trp Tyr Arg Val Val Ala Glu Val Gln Ile Cys His Gly Lys
86   65           70           75           80
88 ACG GAG GCT GTG GGC CAG GTC CAC ATC TTC TTC CAG GAT GGG ATG GTG      288
89 Thr Glu Ala Val Gly Gln Val His Ile Phe Phe Gln Asp Gly Met Val
90           85           90           95
92 ACG TTG ACT CCA AAC AAG GGT GTG TGG GTG AAT GGT CTC CGA GTG GAT      336
93 Thr Leu Thr Pro Asn Lys Gly Val Trp Val Asn Gly Leu Arg Val Asp
94           100          105          110
96 CTC CCA GCT GAG AAG TTA GCA TCT GTG TCC GTG AGT CGT ACA CCT GAT      384
97 Leu Pro Ala Glu Lys Leu Ala Ser Val Ser Val Ser Arg Thr Pro Asp
98           115          120          125
100 GGC TCC CTG CTA GTC CGC CAG AAG GCA GGG GTC CAG GTG TGG CTT GGA      432
101 Gly Ser Leu Leu Val Arg Gln Lys Ala Gly Val Gln Val Trp Leu Gly
102           130          135          140
104 GCC AAT GGG AAG GTG GCT GTG ATT GTG AGC AAT GAC CAT GCT GGG AAA      480
105 Ala Asn Gly Lys Val Ala Val Ile Val Ser Asn Asp His Ala Gly Lys
106 145           150          155          160
108 CTG TGT GGG GGC CTK TGG AAA ATTTGACGGG GGACCAGACC AATGATTGGG      531
W--> 109 Leu Cys Gly Gly Xaa Trp Lys
110           165
112 ATGATTCCCA GGAGAAGCCA GCGATTGGGG AAWTGGAGAG CGCAGGGACT TTCTYCCMCA      591
114 TGTTAATGGG CTTGWTCCAG TTCATCCCAC CAGGAACGAA GGATTTT      638
117 (2) INFORMATION FOR SEQ ID NO: 2:
119     (i) SEQUENCE CHARACTERISTICS:
120         (A) LENGTH: 167 amino acids
121         (B) TYPE: amino acid
122         (D) TOPOLOGY: linear
124     (ii) MOLECULE TYPE: protein
126     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
W--> 128 Ala Arg Gln Leu Ala Ala Xaa Gln Ala Val Tyr Val Lys Val Lys Ala
129   1           5           10           15
131 Glu Ala Arg Glu Leu Leu Gly His Pro Trp Ser Leu Cys Pro Val Cys
132           20           25           30
134 Gly Cys Gln Leu Thr Thr Phe Asp Gly Ala Arg Gly Ala Thr Thr Leu
135           35           40           45
137 Leu Val Ser Met Lys Leu Ser Ser Arg Cys Pro Gly Leu Gln Asn Thr

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/988,292

DATE: 01/15/2002

TIME: 18:58:37

Input Set : N:\Crf3\RULE60\09988292.raw

Output Set: N:\CRF3\01152002\I988292.raw

```

138      50      55      60
140 Ile Pro Trp Tyr Arg Val Val Ala Glu Val Gln Ile Cys His Gly Lys
141 65      70      75      80
143 Thr Glu Ala Val Gly Gln Val His Ile Phe Phe Gln Asp Gly Met Val
144      85      90      95
146 Thr Leu Thr Pro Asn Lys Gly Val Trp Val Asn Gly Leu Arg Val Asp
147      100      105      110
149 Leu Pro Ala Glu Lys Leu Ala Ser Val Ser Val Ser Arg Thr Pro Asp
150      115      120      125
152 Gly Ser Leu Leu Val Arg Gln Lys Ala Gly Val Gln Val Trp Leu Gly
153      130      135      140
155 Ala Asn Gly Lys Val Ala Val Ile Val Ser Asn Asp His Ala Gly Lys
156 145      150      155      160
W--> 158 Leu Cys Gly Gly Xaa Trp Lys
159      165
161 (2) INFORMATION FOR SEQ ID NO: 3:
163 (i) SEQUENCE CHARACTERISTICS:
164 (A) LENGTH: 874 base pairs
165 (B) TYPE: nucleic acid
166 (C) STRANDEDNESS: single
167 (D) TOPOLOGY: linear
169 (ii) MOLECULE TYPE: cDNA
172 (ix) FEATURE:
173 (A) NAME/KEY: CDS
174 (B) LOCATION: 1..705
176 (ix) FEATURE:
177 (A) NAME/KEY: mat_peptide
178 (B) LOCATION: 1..705
181 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
183 CAG GAC TGC GTG TGC ACG GAC AAG GTG GAC AAC AAC ACC CTG CTC AAC      48
184 Gln Asp Cys Val Cys Thr Asp Lys Val Asp Asn Asn Thr Leu Leu Asn
185 1      5      10      15
187 GTC ATC GCC TGC ACC CAC GTG CCC TGC AAC ACC TCC TGC AGC CCT GGG      96
188 Val Ile Ala Cys Thr His Val Pro Cys Asn Thr Ser Cys Ser Pro Gly
189      20      25      30
191 TTC GAA CTC ATG GAG GCC CCC GGG GAG TGC TGT AAG AAG TGT GAA CAG      144
192 Phe Glu Leu Met Glu Ala Pro Gly Glu Cys Cys Lys Lys Cys Glu Gln
193      35      40      45
195 ACG CAC TGT ATC ATC AAA CGG CCC GAC AAC CAG CAC GTC ATC CTG AAG      192
196 Thr His Cys Ile Ile Lys Arg Pro Asp Asn Gln His Val Ile Leu Lys
197      50      55      60
199 CCC GGG GAC TTC AAG AGC GAC CCG AAG AAC AAC TGC ACA TTC TTC AGC      240
200 Pro Gly Asp Phe Lys Ser Asp Pro Lys Asn Asn Cys Thr Phe Phe Ser
201 65      70      75      80
203 TGC GTG AAG ATC CAC AAC CAG CTC ATC TCG TCC GTT TCC AAC ATC ACC      288
204 Cys Val Lys Ile His Asn Gln Leu Ile Ser Ser Val Ser Asn Ile Thr
205      85      90      95
207 TGC CCC AAC TTT GAT GCC AGC ATT TGC ATC CCG GGC TCC ATC ACA TTC      336
208 Cys Pro Asn Phe Asp Ala Ser Ile Cys Ile Pro Gly Ser Ile Thr Phe

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/988,292

DATE: 01/15/2002

TIME: 18:58:37

Input Set : N:\Crf3\RULE60\09988292.raw

Output Set: N:\CRF3\01152002\I988292.raw

```

209          100          105          110
211 ATG CCC AAT GGA TGC TGC AAG ACC TGC ACC CCT CGC AAT GAG ACC AGG      384
212 Met Pro Asn Gly Cys Cys Lys Thr Cys Thr Pro Arg Asn Glu Thr Arg
213          115          120          125
215 GTG CCC TGC TCC ACC GTC CCC GTC ACC ACG GAG GTT TCG TAC GCC GGC      432
216 Val Pro Cys Ser Thr Val Pro Val Thr Thr Glu Val Ser Tyr Ala Gly
217          130          135          140
219 TGC ACC AAG ACC GTC CTC ATG AAT CAT TGC TCC GGG TCC TGC GGG ACA      480
220 Cys Thr Lys Thr Val Leu Met Asn His Cys Ser Gly Ser Cys Gly Thr
221 145          150          155          160
223 TTT GTC ATG TAC TCG GCC AAG GCC CAG GCC CTG GAC CAC AGC TGC TCC      528
224 Phe Val Met Tyr Ser Ala Lys Ala Gln Ala Leu Asp His Ser Cys Ser
225          165          170          175
227 TGC TGC AAA GAG GAG AAA ACC AGC CAG CGT GAG GTG GTC CTG AGC TGC      576
228 Cys Cys Lys Glu Glu Lys Thr Ser Gln Arg Glu Val Val Leu Ser Cys
229          180          185          190
231 CCC AAT GGC GGC TCG CTG ACA CAC ACC TAC ACC CAC ATC GAG AGC TGC      624
232 Pro Asn Gly Gly Ser Leu Thr His Thr Tyr Thr His Ile Glu Ser Cys
233          195          200          205
235 CAG TGC CAG GAC ACC GTC TGC GGG CTC CCC ACC GGC ACC TCC CGC CGG      672
236 Gln Cys Gln Asp Thr Val Cys Gly Leu Pro Thr Gly Thr Ser Arg Arg
237          210          215          220
239 GCC CGG CGT TCC CCT AGG CAT CTG GGG AGC GGG TGAGCGGGGT GGGCACAGCC      725
240 Ala Arg Arg Ser Pro Arg His Leu Gly Ser Gly
241 225          230          235
243 CCTTCACTGC CCTCGACAGC TTTACCTCCC CCGGACCCTC TGAGCCTCCT AAGCTCGGCT      785
245 TCCTCTCTTC AGATATTTAT TGTCTGAGTT TTTGTTCAGT CCTTGCTTTC CAATAATAAA      845
247 CTCAGGGGGA CATGCAAAAA AAAAAAAAAA      874
250 (2) INFORMATION FOR SEQ ID NO: 4:
252     (i) SEQUENCE CHARACTERISTICS:
253         (A) LENGTH: 235 amino acids
254         (B) TYPE: amino acid
255         (D) TOPOLOGY: linear
257     (ii) MOLECULE TYPE: protein
259     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
261 Gln Asp Cys Val Cys Thr Asp Lys Val Asp Asn Asn Thr Leu Leu Asn
262 1          5          10          15
264 Val Ile Ala Cys Thr His Val Pro Cys Asn Thr Ser Cys Ser Pro Gly
265          20          25          30
267 Phe Glu Leu Met Glu Ala Pro Gly Glu Cys Cys Lys Lys Cys Glu Gln
268          35          40          45
270 Thr His Cys Ile Ile Lys Arg Pro Asp Asn Gln His Val Ile Leu Lys
271          50          55          60
273 Pro Gly Asp Phe Lys Ser Asp Pro Lys Asn Asn Cys Thr Phe Phe Ser
274 65          70          75          80
276 Cys Val Lys Ile His Asn Gln Leu Ile Ser Ser Val Ser Asn Ile Thr
277          85          90          95
279 Cys Pro Asn Phe Asp Ala Ser Ile Cys Ile Pro Gly Ser Ile Thr Phe
280          100          105          110

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/988,292

DATE: 01/15/2002

TIME: 18:58:37

Input Set : N:\Crf3\RULE60\09988292.raw

Output Set: N:\CRF3\01152002\I988292.raw

282 Met Pro Asn Gly Cys Cys Lys Thr Cys Thr Pro Arg Asn Glu Thr Arg
 283 115 120 125
 285 Val Pro Cys Ser Thr Val Pro Val Thr Thr Glu Val Ser Tyr Ala Gly
 286 130 135 140
 288 Cys Thr Lys Thr Val Leu Met Asn His Cys Ser Gly Ser Cys Gly Thr
 289 145 150 155 160
 291 Phe Val Met Tyr Ser Ala Lys Ala Gln Ala Leu Asp His Ser Cys Ser
 292 165 170 175
 294 Cys Cys Lys Glu Glu Lys Thr Ser Gln Arg Glu Val Val Leu Ser Cys
 295 180 185 190
 297 Pro Asn Gly Gly Ser Leu Thr His Thr Tyr Thr His Ile Glu Ser Cys
 298 195 200 205
 300 Gln Cys Gln Asp Thr Val Cys Gly Leu Pro Thr Gly Thr Ser Arg Arg
 301 210 215 220
 303 Ala Arg Arg Ser Pro Arg His Leu Gly Ser Gly
 304 225 230 235

306 (2) INFORMATION FOR SEQ ID NO: 5:

308 (i) SEQUENCE CHARACTERISTICS:

309 (A) LENGTH: 1209 base pairs

310 (B) TYPE: nucleic acid

311 (C) STRANDEDNESS: single

312 (D) TOPOLOGY: linear

314 (ii) MOLECULE TYPE: cDNA

319 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

321 ATTGGTGCTA CCTGGCTCTC CTGTCTCTGC AGCTCTACAG GTGAGGCCCA GCAGAGGGAG 60
 323 TAGGGCTCGC CATGTTTCTG GTGAGCCAAT TTGGCTGATC TTGGGTGTCT GAACAGCTAT 120
 325 TGGGTCCACC CCAGTCCCTT TCAGCTGCTG CTTAATGCCC TGCTCTCTCC CTGGCCCACC 180
 327 TTATAGAGAG CCCAAAGAGC TCCTGTAAGA GGGAGAACTC TATCTGTGGT TTATAATCTT 240
 329 GCACGAGGCA CCAGAAGTCT CCCTGGGTCT TGTGAATGAA CTACATTTAT CCCCTTTCCT 300
 331 GCCCCAACCA CAAACTCTTT CCTTCAAAGA GGGCTGCCT GGTTCCTCC ACCCAACTGC 360
 333 ACCATGAGAT CGGTCCAAGA GTCCATTCCC CAGGTGGGAG CCAACTGTCA GGGAGGTCTT 420
 335 TCCCACCAAA CATCTTTCAG TTGCTGGGAG GTGACCATAG GGCTCTGCTT TTAAAGATAT 480
 337 GGCTGCTTCA AAGGCCAGAG TCACAGGAAG GACTTCTTCC AGGGAGATTA GTGGTGATGG 540
 339 AGAGGAGAGT TAAAATGACC TCATGTCCTT CTTGTCCACG GTTTTGTTGA GTTTTCACTC 600
 341 TTCTAATGCA AGGGTCTCAC ACTGTGAACC ACTTAGGATG TGATCACTTT CAGGTGGCCA 660
 343 GGAATGTTGA ATGTCTTTGG CTCAGTTCAT CTAAAAAAGA TATCTATTG AAAGTTCTCA 720
 345 GAGTTGTACA TATGTTTCAC AGTACAGGAT CTGTACATAA AAGTTTCTTT CCTAAACCAT 780
 347 TCACCAAGAG CCAATATCTA GGCATTTTCT CCGTAGCACA AATTTTCTNA TTGCTTAGAA 840
 349 AATTGTCCTC CCTGTTCTTT CTGTCTGNAG ACTTAAGTGA GTTAGGTCTT TAAGGAAAGC 900
 351 AACGCTCCTC TGAAGTGCTT GTCTTTTTTC TGTGCGGAA ATAGCTGGTC CTTTTTCGGG 960
 353 AGTTAGATGT ATAGAGTGTT TGTATGTAAA CATTTCTTGT AGGCATCACC ATGAACANAG 1020
 355 ATATATTTTC TATTANTTA NTATATGTGC ACTTCAAGAA GTCAGTGTCA GAGAAATAAA 1080
 357 GAATTGTCTT AAATGTCATG ATTGGAGATG TCCTTTGCAT TGCTTGGAAG GGGGTGTACCT 1140
 359 AGAGCCAAGG AAATTGGCTC TGGTTTGGAA AAATTTTGCT GTTATTATAG TAAACATACA 1200
 361 AAGGATGTC 1209

363 (2) INFORMATION FOR SEQ ID NO: 6:

365 (i) SEQUENCE CHARACTERISTICS:

366 (A) LENGTH: 548 base pairs

367 (B) TYPE: nucleic acid

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/988,292

DATE: 01/15/2002

TIME: 18:58:38

Input Set : N:\Crf3\RULE60\09988292.raw

Output Set: N:\CRF3\01152002\I988292.raw

L:28 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:29 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:69 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:109 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:128 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:158 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2